

⑫ 公開特許公報 (A) 昭61-143799

⑬ Int.Cl.⁴

G 10 L 5/02

識別記号

庁内整理番号

7350-5D

⑭ 公開 昭和61年(1986)7月1日

審査請求 有 発明の数 1 (全8頁)

⑮ 発明の名称 音声合成方式

⑯ 特願 昭60-220060

⑰ 出願 昭54(1979)7月13日

⑱ 特願 昭54-88120の分割

⑲ 発明者 小松 昭男 国分寺市東恋ヶ窪1丁目280番地 株式会社日立製作所中央研究所内

⑳ 出願人 株式会社日立製作所 東京都千代田区神田駿河台4丁目6番地

㉑ 代理人 弁理士 小川 勝男 外1名

明細書

発明の名称 音声合成方式

特許請求の範囲

1. 音節を合成するための情報をメモリに記憶しておき、合成すべき音節を指示することにより、上記メモリから所望の情報を読み出し音声波形を合成してなる音声合成方式において、上記音節の音程を指示する手段と、上記音節の長さを指示する手段と、上記音節の大きさを指示する手段とを有し、これらの手段の指示に従って、上記メモリから読み出される上記所望情報を制御することにより上記音声波形を合成することを特徴とする音声合成方式。

発明の詳細な説明

〔発明の利用分野〕

本発明は、所望の音声を合成して出力する音声合成方式、特に、日本語を合成して出力するのに適した音声合成方式に関するものである。

〔発明の背景〕

近年、マン・マシン・インターフェースとして、

小型端末や、さらには家電品や自動車にまで、言葉をしゃべらせたいという要求が高まってきている。

これらの要求は、それぞれの用途によって、たとえば、キーインされた文章を復唱することであったり、「シートベルトをしめて下さい」等の警告を状況によって選択して発声することであったりする。したがって、無限の語彙を合成できる方式が望まれる。

これに対し、従来の音声合成方式としては、あらかじめ決められた単語、あるいは文章を記憶しておき、これをそのままあるいは組み合わせたりしながら、発声するという方式がある（例えば特開昭53-116705参照）。この中でも音声波形をそのまま記憶する方式や、これを一度処理して、情報量を削減して記憶する方式などがある。

前者は、莫大な記憶容量を必要とするため、小型の機器には適用できない。

後者では、パーコール（PARCOR）方式と呼ばれるものが、現在最も有望視されている。こ

の方式では、あらかじめ用意した単語や文章に対して PARCOR 分析という処理を行ない、これによって、音声合成に必要な情報 (PARCOR 係数と呼ぶ。) を抽出し、これを合成フィルタの入力として、音声を合成する。それ故、音声を記憶するためのメモリ容量を、かなり削減することができる。

しかしながら、PARCOR 方式では、用途が変わることに、必要な単語や文章を用意し、PARCOR 分析を行なうという、めんどうな前処理が必要となる。

さらに、用意すべき文章の量に対応して、PARCOR 係数を記憶すべきメモリも増加するという欠点を持っている。

〔発明の目的〕

本発明の主な目的は、一定の小容量のメモリを用いるだけで、無限の語彙を合成できる音声合成方式を提供することにある。

本発明の他の目的は、特殊な前処理を必要とすることなく、音声を合成できる音声合成方式を提

供することにある。

本発明のさらに他の目的は、極めて自然な音声を合成できる音声合成方式を提供することにある。

本発明のさらに他の目的は、汎用のマイクロコンピュータのような、単純な制御回路で音声を合成できる音声合成方式を提供することにある。

〔発明の概要〕

このような目的を達成するために、本発明では、基本的には、かな 1 文字が 1 音節に対応し、この音節は子音部と母音部とに分解できるという、日本語の特徴に着目し、少數の子音音素の波形と母音音素の波形とをメモリに記憶しておき、指示入力によりこのメモリから所望の子音音素と母音音素との波形を読み出して、これらの波形を連結することによって音声を合成したことに特徴がある。

また、本発明では、鼻音性子音音素 (後で定義する。) を有する音節については、子音音素の波形から母音音素の波形への遷移部が聴覚上重要な要素になることに鑑みて、鼻音性子音音素の波形に母音音素の波形を連結する場合に、両音素の波

形のピッチを合わせるとともに、両音素の波形を重みつき加算平均して得られる波形を両音素の波形の間に挿入して補間を行ない、自然な音声の合成を可能にしたことに特徴がある。

〔発明の実施例〕

以下、本発明を実施例により詳細に説明する。第 1 図は本発明による音声合成方式を実現する装置の一実施例のブロック図を示すものである。

図において、1 は音声データメモリで、単音節ごとに抽出した子音音素の波形と母音音素の波形がデジタル化して記憶されている。2 は 1 チップ・マイクロコンピュータ等からなる処理装置で各種の演算・制御を行なう中央制御部 2 1 と、発音プログラムを格納する読み出し専用メモリ (ROM) 2 2 と、プログラム実行時に必要となるデータの一格納用の読み書き可能なメモリ (RAM) 2 3 と、各種のクロックを発生するクロック発生器 2 4 とから主に構成されている。

3 は処理装置 2 で得られたデジタルのデータをアナログ波形に変換するデジタル/アナログ

変換器 (DA 変換器)、4 は DA 変換器 3 の出力を増幅する増幅器、5 は音声を出力するスピーカ、6 は合成すべき音声に相当する文字列を入力する端子である。

かな文字に対応する各音節は、子音部と母音部とに分離できるから、これを分りやすくするために、ローマ字で、各音節を KA, SI など書き、子音部を構成する子音音素を K, S、母音部を構成する母音音素を A, I などと書くことにする。

すると、各音節を構成する母音部は A, I, U, E, O の 5 つの母音音素の 1 つからなり、子音部は K, S, T, N, H, M, Y, R, W, G, Z, D, B, P の 14 個の子音音素の 1 つからなっていることになり、これら 5 個の母音音素の波形と、14 個の子音音素の波形と、さらに、特殊音である「ん」に相当する子音音素 NN の波形とを音声データメモリ 1 に記憶しておく。

これらの波形を音声データメモリ 1 に記憶するために、人間が発する、かな文字の音声をマイクロホンで電気信号に変換し、一定の周期、例えば、

10 KHzでサンプリングして、その振幅をデジタル化して、例えば、8ビットのデータを得、そのようにして得られたサンプリングデータを部分的に選択して対応する音素の波形としている。

そして、音声を合成して出力するときには、入力端子6からのかな文字の指示により、まず、子音音素に対応する波形をメモリ1から読み出して出力した後、続けて母音音素に対応する波形を同じくメモリ1から読み出して出力するようすれば、日本語の50音とすべての濁音とが合成できる。具体的には、入力端子6から文字列が入力されると、処理装置2内の中央制御部21ではROM22内のプログラムに従って、対応する子音音素および母音音素の波形を音声データメモリ1から読み出し、所定の処理を施した後、順次DA変換器3に出力する。例えば、「わたくしは……」という文字列が入力されると、まず、音素W、続けてA、さらに、T、A、K、U、……を出力する。

このようにすれば、5つの母音音素と、14個

第2図は代表的な音声波形を模式的に示したもので、(a)は音節A、(b)は音節SA、(c)は音節NA、(d)は音節GAの波形を示している。

この図から解るように、鼻音性の子音音素(例えば、N)と母音音素(例えばA)との音声波形は明らかなピッチを持っている。そこで、本発明では、子音音素と母音音素との波形のピッチを合わせて両者を連続するようにしている。すなわち、本発明では、子音音素および母音音素の波形のすべての部分をデータメモリ1に記憶するのではなく、適当なピッチ間隔をおいた、1ピッチ分ずつの波形を1~数個選択し、これらを記憶している。例えば、第2図においては、破線で囲まれた波形 α 、 β 、 α' 、 β' 、 γ などと選んで記憶する。

また、上述した遷移部での不連続性が感知されるのは、子音音素から突然、波形レベルの全く異なる母音音素に接続されるためであることに鑑みて、本発明では、この遷移部において、波形レベルが滑らかに変化するように一種の補間法を用い

の子音音素と、特殊な1個の子音音素とを記憶するだけで、すべての音節を合成できるので、予じめ音節を記憶しているものに比べても約1/5のメモリ容量でよいことになり、極めて少ないメモリ容量ですべての語彙の合成が可能となる。

また、一旦音声データメモリ1に書き込まれると、用途などの変更によってデータメモリの内容を書き換えるなどの必要もなく、面倒な前処理が省略できる。

ところで、本発明者らによる実験結果では、子音音素の内N、M、Y、R、Wの音素(これらの音素は一般には雑音を伴なわない有性子音音素と言われているが、ここでは便宜上、鼻音性子音音素という。)を音声合成した結果、得られた音声出力は自然感に欠けており、その原因は鼻音性子音音素から母音音素への遷移部が連続になるからであることが解った。

そこで、本発明の実施例では、このような問題を解決するために、以下に示すような方式を探った。

たことに特徴がある。すなわち、鼻音性子音音素の最後尾の1ピッチ分の波形と、母音音素の最前部の1ピッチ分の波形などを基に、これらを、例えば、3:1、1:1、1:3というように、異なる重みをつけて加算平均した1ピッチ分の波形を数種作って、子音音素の最後部の波形と母音音素の最前部の波形との間に挿入する。

このようにすれば、出力される合成音声波形は子音から徐々に母音に変化するから、人間が聞いた時の不自然感を極度に減少することができる。

また、上述したように、データメモリ1には、適当なピッチ間隔をおいて、選択した1ピッチ分の音声波形を1~数個抽出して記憶しているが、これらのデータを読み出して音声合成を行なう場合、読み出された1ピッチの音声波形を数個ずつ繰り返して出力するとともに、隣接する選択音声波形間を、両音声波形を上述したように重みつき加算平均して得られる数ピッチ分の音声波形で連結する。

以上述べた処理を行なうには、第1図の処理裝

置2において、音声データメモリ1から読み出した2種の音声データ、すなわち、子音音素の波形と母音音素の波形との組み合わせまたは、同一音素内の2つの波形を読み出し、両者を重みつき加算平均した後、DA変換器3に出力すればよい。

また、ま子音音素の内、濁音性子音音素G, Z, D, Bの波形を上述したように、単に母音音素の波形と結合するだけでは自然感に欠ける合成音声となるが、濁音性の音節は、実際には、K, S, T, Hなどの雑音状の子音音素の波形で始まり、次いで、鼻音性の子音音素の波形が表われ、最後に母音音素の波形に移行している。例えば、第2図(d)のように、音節G Aは、第2図(b)と同じように、雑音状の子音音素の波形で始まり、次いで、第2図(c)と同じように、鼻音性の子音音素の波形が表われ、最後に、第2図(a)と同じように母音音素の波形に移行する。

したがって、濁音性音節の合成を行なう場合は、最初に雑音状子音音素の波形を出力し、次いで、鼻音性子音音素の波形を出力し、次に、前述した

入力文字が雑音状の子音音素を指定している場合には、この雑音状子音音素の波形を出力した後、母音音素のときと同様の音声波形出力ルーチンを経て、文字入力ルーチンに戻る。

さて、入力文字が鼻音性の子音音素を指定している場合には、処理装置2のプログラムは、前述の補間による合成を実行する。今、音節「な」(N A)を合成するものとする。音声データメモリ1には、前述したように、第2図中、鼻音性の子音音素に相当する1ピッチ分の音声波形 α , β と、母音音素の波形 γ が予じめ用意されており、それらの情報が順次読み出される。

処理装置2は、第3図の処理手順により、まず始めに、波形 α を4回繰り返して出力する。次に、波形 α と β を重みつき加算平均した補間波形を3種類出力し、続いて波形 β を出力し、さらに波形 β と γ についても同様に補間波形を3種類出力した後、母音音素の波形 γ を4回繰り返して文字入力ルーチンに戻る。第3図において、たとえば $\frac{3}{4}\alpha + \frac{1}{4}\beta$ と書いたのは、 α を $\frac{3}{4}$ 倍したものに β

重みつき加算による補間を行なった後、母音音素の波形を出力すればよい。

なお、雑音状音素の波形と半濁音音素Pの波形は、上述したような処理を施さなくとも、母音音素の波形と単に結合することにより自然な合成音が得られた。

第3図および第4図は第1図の処理装置2で実行される動作の一例のフローチャートを示すもので、第3図は音声合成のフローチャート、第4図は音声波形出力のフローチャートを示す。

以下、処理装置2の動作を第3図、第4図により詳細に説明する。

処理装置2は、電源が投入されると、文字入力の待ち受け状態となる。そして、入力端子6から一文字が入力されると、まず第3図に示すように、対応する音素の種類を判定する。入力文字が母音音素を指定している場合には、その母音音素の1ピッチ分の波形(例えば、第2図の γ)を8回繰り返して出力した後、次の文字を受け取るルーチンに戻る。

を $\frac{1}{4}$ 倍したものを加えて得られる波形という意味である。

一方、入力文字が濁音性音素を指定している場合には、始めに雑音状子音音素の波形を出力した後、鼻音性音節と同様の手順で音声合成を行なうのであるが、実験の結果、母音音素の波形 γ を使用しないで、濁音音素の音声波形の後部から採取した波形 β' の繰り返しで、合成を終える方がより自然な感じが得られることが分ったので、そのようにしている。

なお、以上の説明で用いた数値例は、男声データに対して自然な感じの合成音が得られるよう選んだ1例である。たとえば女声の場合は1ピッチの持続時間が短いので、この繰返し回数を増して、音節の長さを適切に調整する必要がある。

さて次に、今述べた音声波形出力ルーチンについて、詳細に説明する。このルーチンは、第4図に示すように、音声データメモリ中に格納されたデジタル波形を、ポインタを進めながら、逐次読み出して、一定の周期(この例では $100\mu s$)

で、DA変換器3へ出力するようになっている。また、第3図のメインルーチンから呼ばれるサブルーチンの形式をとっており、鼻音性子音音素の波形を出力する場合には、ピントを2つ用いて、 α 、 β などの2つの波形を読み出し、それらの波形の重みつき加算を、リアルタイムで実行できるようにしてある。ここで重みを $\frac{1}{2}$ 、 $\frac{1}{4}$ というように、 2^{-n} の形にしておくと、シフト命令だけで分数計算が行なえるので、処理装置2のプログラムに適した設計となる。

なお、第4図中、「100μs待つ」と書いているが、実際には、前後のプログラムステップに必要な時間を差引いた量の待ち時間を設けることによってAD変換器への出力が、100μs毎に行なわれるようしている。

なお、上述した実施例では、入力端子6から所定望の音節を指定する例について述べたが、その音節情報とともに種々の情報を付加して入力することにより、以下に述べるように特殊な音声を合成できる。

コードに応じて、読み出された音素の1ピッチ分の波形の後部を切り捨て、その波形の長さを制御して出力する。

また、本発明では、1ピッチ分ずつの音声波形を何回か繰り返して出力することによって音節を得ているが、この繰り返し回数を変えることによって音節の長さを変えることができる。そこで、本発明の実施例では、音節情報の他に、第5図(b)に示すような、音節の長さに対応するコード、例えば、数字を付加して処理装置に入力し、入力されたコードに応じて1ピッチ分の繰り返し回数を変えることによって、音節の長さを制御できる。

さらに、本発明の実施例では、音節情報の他に、音の大きさに対応するコードを付加して入力し、それに応じて、音素波形の振幅を変えることにより、音の大きさを任意に制御できる。

このように、入力端子6から音節を指定とともに、音の高さ、長さ、大きさを指定し、それに応じて、選択された音節の高さ、長さ、大きさ

前述したように、母音音素と鼻音性子音音素の波形は明らかなピッチを持った周期性の波形であるので、音声データメモリ1には、1ピッチ分ずつの波形が記憶されることになる。そこで、この1ピッチ分の長さの波形をすべて出力するのではなく、途中までの波形を繰り返し出力すると、ちょうど声の高さ(音程)が高くなつたように聞こえる。また、雑音状の子音音素の波形は周期性を持たないので、この後に連結される母音のピッチによって、その音節の音程が決定される。したがって、合成音声のすべての音節に音程を与えることができる。なお、この場合、波形を大巾に切り捨てて、極度に高い音程にしない限り、了解度が悪化することはない。

そこで、本発明の実施例では、処理装置2に文字列を入力する時に、各かな文字を構成する音節情報の他に、第5図(a)に示すような、各音程と対応するコード、例えば、英数字を付加し、それによって音程を指定するようにした。このような指定がなされると、処理装置2では、英数字の

を制御することにより、合成音声のアクセントを変えたり、音声を発生させることができる。

第6図は本発明による音声合成方式を実現する装置の他の実施例を示すもので、家電品や自動車において、警告等を発生させるために使用される。

第6図において、第1図と異なる点は、検出器7からの検出信号を処理装置2に入力していることと、音声データメモリ1以外に、文章データメモリ8が付加されていることである。

この文章データメモリ8には、単語あるいは文章などの文字列が記憶されている。

ここで、自動車における半ドアに対して警告を発する場合を例にとると、検出器7としては半ドア検出器を使用し、文章データメモリ8には「半ドアです」という文章コードが記憶されている。

処理装置2は、通常、他の仕事、例えば、ユーザからのキーイン、温度、エンジン回転数などを入力して、これらに応じた処理を行なっているが、検出器7により、半ドアが検出されると、前述した音声合成プログラムを起動する。音声合成プロ

グラムにより、必要な文章、すなわち、「半ドアです。」という文章を文章データメモリ8から順次読み出して、これに応じて、前述したと同様にして音声を合成し、スピーカ5から警告として出力する。

上述した例では、自動車の半ドアの警告の場合について述べたが、種々の応用が考えられることは言うまでもない。

また、文章データメモリ8に種々の文章を記憶しておき、対応する種々の検出器の出力により、所望の文章を選択して発声せしめるようにしてもよい。

また、検出器の出力を用いないで、人間の指示入力により、特定の文章をデータメモリ8から読み出して、対応する音声を合成するようにしてもよい。

なお、上述した実施例では、マイクロコンピュータ等の処理装置を使用する例について説明したが、本発明は、それに限定されるものではなく、上述した手法で音声合成できる回路であれば、ど

のものでもよい。

また、子音音素としては、上述したもの以外に、例えば、外来語で用いるF, Lなどを追加することも可能である。

【発明の効果】

以上述べたように、本発明によれば、一定のメモリ容量で、言い換えれば、非常にコンパクトに、無限の語彙を合成できる装置が実現できる。

また、PARCOR方式のような特殊な前処理は必要ではなく、用途が変る毎に、新しい音声データを用意するための前処理も必要でない。

さらに、音素の波形の間の補間を行なうことにより、極めて自然な音声を合成できる。

さらに、音節情報に、音の高さ、強弱、長さなどの情報を附加して指示することにより、音にアクセントをつけたり、歌声を発生させたりすることができる。

さらに、汎用のマイクロコンピュータを用いて容易に実現できるので、このマイクロコンピュータの空き時間を他の用途に共用すれば、極めて合

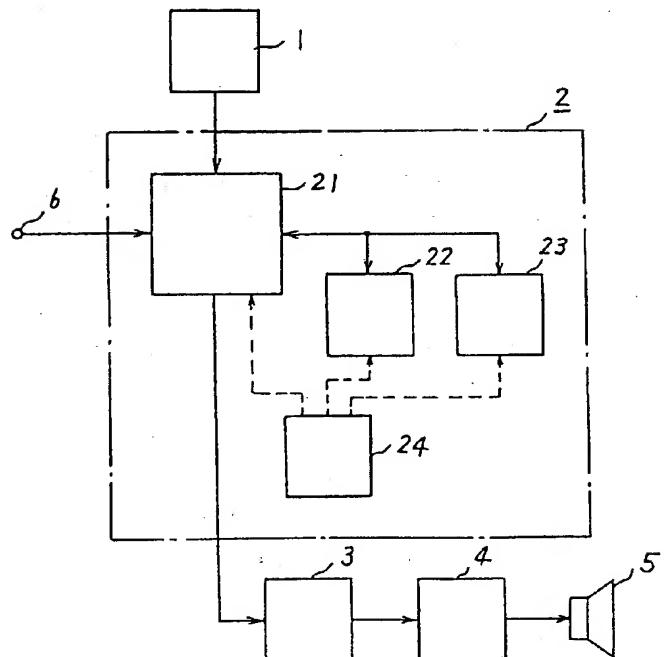
理的な、端末機器、家電品、などを構成できる。さらに、低コスト性を生かして、各種の教育機器、おもちゃなどにも応用できる。

図面の簡単な説明

第1図は本発明による音声合成方式を実現する装置の一実施例のブロック図、第2図は代表的な音声波形図、第3図および第4図は第1図の処理装置における音声合成および音声波形出力の処理手順の一例を示すフローチャート、第5図は音節情報に附加して入力される情報コードを示す図、第6図は本発明による音声合成方式を実現する装置の他の実施例のブロック図である。

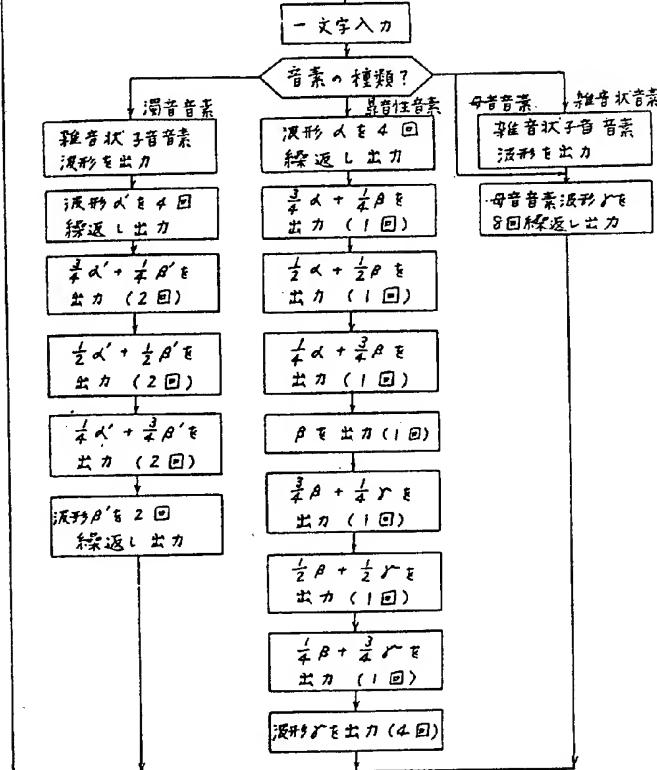
1：音声データメモリ、2：処理装置、3：スピーカ、6：入力端子。

第1図

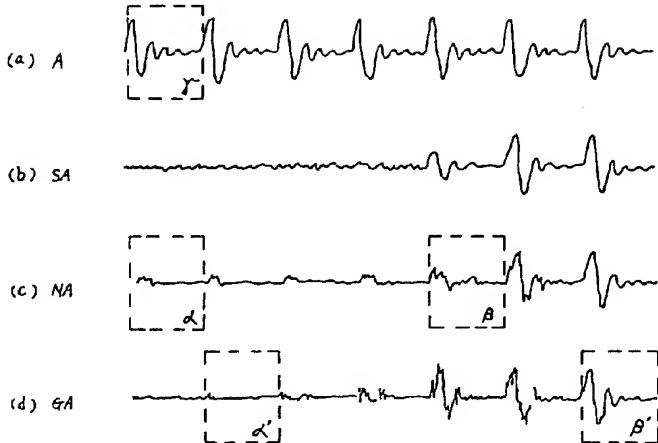


代理人弁理士小川勝男

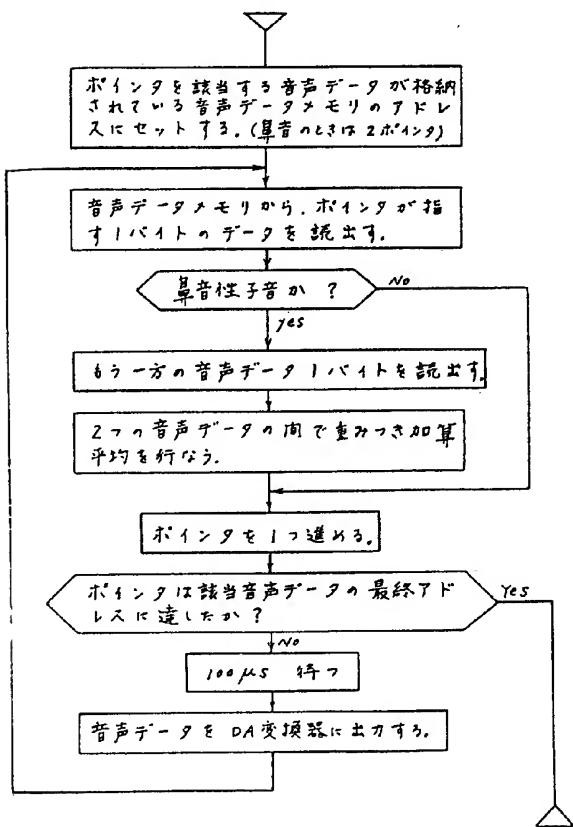
第3図



第2図



第4図



第5図

(A)

英数字	音程
0	八
1	嬰八
2	二
3	嬰二
4	六
5	ヘ
6	嬰ヘ
7	ト
8	嬰ト
9	イ
A	嬰イ
B	口
C	ハ

(B)

数字	長さ
0	全音符
1	2分音符
2	4 "
3	8 "
4	16 "
5	32 "
6	付夷2分音符
7	" 4 "
8	" 8 "
9	" 16 "

第 6 図

